April 2018

# Development of Multilingual Resource Management Mechanisms for Libraries

Sukumar Mandal
*Department of Library and Information Science, The University of Burdwan*, sukumar.mandal5@gmail.com

# Development of Multilingual Resource Management Mechanisms for Libraries

Dr. Sukumar Mandal
Assistant Professor, Department of Library and Information Science
The University of Burdwan, Burdwan – 713 104
Email: sukumar.mandal5@gmail.com

## Abstract

Multilingual is one of the important concept in any library. This study is create on the basis of global recommendations and local requirement for each and every libraries. Select the multilingual components for setting up the multilingual cluster in different libraries to each user. Development of multilingual environment for accessing and retrieving the library resources among the users as well as library professionals. Now, the *methodology of integration of Google Indic Transliteration* for libraries have follow the five steps such as (i) selection of transliteration tools for libraries (ii) comparison of tools for libraries (iii) integration Methods in Koha for libraries (iv) Development of Google indic transliteration in Koha for users (v) testing for libraries (vi) results for libraries. Development of multilingual framework for libraries is also an important task in integrated library system and in this section have follow the some important steps such as (i) Bengali Language Installation in Koha for libraries (ii) Settings Multilingual System Preferences in Koha for libraries (iii) Translate the Modules for libraries (iv) Bengali Interface in Koha for libraries. Apart from these it has also shows the Bengali data entry process in Koha for libraries such as Data Entry through Ibus Avro Phonetics for libraries and Data Entry through Virtual Keyboard for libraries. Development of Multilingual Digital Resource Management for libraries by using the DSpace and Greenstone. Management of multilingual for libraries in different areas such as federated searching (VuFind Multilingual Discovery tool ; Multilingual Retrieval in OAI-PMH tool ; Multilingual Data Import through Z39.50 Server ). Multilingual bibliographic data edit through MarcEditor for the better management of integrated library management system. It has also create and editing the content by using the content management system tool for efficient and effective retrieval of multilingual digital content resources among the users.

**Keywords :** Google Indic Transliteration, Koha, DSpace, Greenstone, Bengali Avro Keyboard, SCIM, Federated searching tool, and MarcEditor

## 1.0 Introduction

Development of multilingual in domain specific cluster in one of the important tasks for two purposes housekeeping operations and information retrieval system to the users as well as librarians. Most of the college libraries are suffering to managed their multilingual documents and users want to Bengali language in different subject areas including Bengali, physics, chemistry, geography, history and etc. In most of the library management softwares not supported the multilingual documents but this research work tries to solved this problem through Koha. Now, in college libraries users are finding their necessary documents in the bibliographic descriptions including author, title, subject and others fields in Bengali language. This will helpful for the users in the libraries. In the last decade, the use of Bengali scripts in daily computer usage has gained wide acceptance in India. Wide ranges of Bengali software have been developed so far to meet the ever-growing demand in the local market (Alshawi, 1992). From the very beginning, Indian software developers followed two different paths. One group started writing software from the scratch, while the other group tried to embed Bengali scripts in popular international software (Angelov, 2008). But it is now well established that due to the limited market size and massive development and upgrading cost involved in writing software from the scratch, embedding Bengali scripts is the most feasible way (Angelov, 2009). This research study focuses primarily on developing a Bengali scripting system capable of sorting Bengali texts linguistically. Although the solution presented in this paper puts no restriction over the method of implementation, we have preferred, for obvious reasons, to embed our solution in Ubuntu interface (Angelov & Ranta, 2009).

Here, in this research paper proved that no completely linguistically sorted Bengali coding scheme exists. We have further proved that it is also not possible to define any rule to derive the complete linguistic order from any partially linguistically ordered Bengali coding scheme (Bar-Hillel, 1964). Based on the nature of the mapping functions, whether any information is lost in transformations or not, two solutions are suggested (Beckert, Hahnle, & Schmitt, 2007). Both of the solutions employ conversion tables to handle the complexity associated with the compound letters. In the second solution we have introduced an internal coding scheme, in addition to the conventional coding scheme, to provide non-lossy transformations (Bender & Flickinger, 2005). This solution gives us some extra benefits (Cook, 1999). Bengali texts, written in a completely unordered coding scheme, can now be sorted. Moreover, based on the fact that non-lossy transformations are reversible, we have developed an application to convert Bengali texts among different coding schemes.

**The European Digital Library (TEL) and the EDL** project generated through the survey of users based on the analysis of log files for user requirements. It is found that weblogs is the search engine where user prepare the own blog and publish it in Internet for access the updated documents by really simple syndication feeds. In this section this problem is to be solved through lifera on Ubuntu operating system. Now a days it is also possible to access the institutional portals by federated search system for the users in college libraries (Janssen, 2003). Translate the documents from the google translator and google input tool in web environment for the document and resources available in Internet or in offline mode (Treble CLEF, 2008). Most of the

users are to be interested in multilingual related documents becasuse they have to studies their own languages from the open source software, open standards and open source tools. Metadata is fundamental to persons, organizations, machines, and an array of enterprises that are increasingly turning to the Web and electronic communication for disseminating and accessing information. Substantiating the growth is the development of metadata schemas supporting proects ranging from restricted corporate websites to freely accessible digital libraries; experimentation with a range of metadata creation tools and techniques; advancements in the development of the semantic web; and an unprecedented developing of diverse communities with a vested interest in resource management and discovery.

**UNESCO (2003) Recommendations**: The idea of multilingual is to be changed from past to present. In modern age peoples are communicate to each other in different languages, in such a way here require the status of language of this World. Most of the peoples is spoken in English language , yet requirements of multilingual concept in databases to display the metedata in the field of digital library. On the other hand also requirements of multilingual bibligraphic and authority information for the college users to access, downloaded the particular resources available in databases. The application is predominantly to the data in motion, objects that users do not physically hold, whose description resides as a part of the object, rather than separately in a library catalogue. Metadata is no longer a new concept. Cataloguers have been employing it as descriptive method for decades as MARC records in OPACs or as card in catalogs. The most innovative aspect of it now is that it has emerged multitude of methods which employ it and the area in which it is being used. TEI, GILS and Dublin Core metadata each comes from a different community or as a collaboration of communities in order to attempt to describe a very slippery publication medium. It is not unlike the chaotic times when printing was first invented. The search is definitely towards an emerging and mutable publication medium for which users have few definitive answers because users have not discovered all of the question yet. As text publishing models increasingly incorporate electronic access and delivery into their paradigm, it becomes clear that metadata becomes included in the editorial decisions involved in the creation of the texts. Thus, this transformation from the old model of simply publishing the text in different languages and leaving the creation of metadata description in the hands of outside agencies, such as libraries or, more specifically, cataloguers. The Greenstone software can be used to serve collections over the World Wide Web. Greenstone can be made available, in precisely the same form, on CD-ROM. The user interface is through a standard web browser (Mozilla) and the interaction is identical to accessing the multilingual collections on the web except that response times are more predictable. Dublin Core metadata element sets is also support the multilingual resource management mechnism. DSpace support html format to manage the multilingual both the admin and user interfaces. Moreover, multilingual concept apply in six basic domain specific cluster to access, download and upload the bibliographic and metadata related information for the users in college libraries. Different search techniques is also applicable in different clusters to manage the multilingual resources in different item types by open source software, open standards and open source tools.

The main objectives of this research paper is as follows:

To designing the framework in Unicode – compliant environment for supporting multilingual document processing and retrieval with special reference to Bengali script for easy implementation in libraries.

### 1.1 Multilingual Components for libraries

Multilingual resource managed through open source tools and standards. There are many standards are available in multilingual for the domain specific cluster in the Libraries. This research paper has select the Unicode based open softwares in six domain specific cluster like integrated library system cluster, digital media archiving cluster, content management system cluster, learning content management system, federated search system and college communication and interaction. The components of multilingual standards are to be represents in the table-1 for designing the multilingual resources in the college libraries.

| | |
|---|---|
| Virtual Keyboard | Bengali, Hindi and Sanskrit |
| Unicode | UTF-8, UTF-16 and UTF-32 |
| Avro Phonetics | Ibus preferences seamles integration |
| SCIM Input methos setup | Run by terminal in Ubuntu |
| L10N | ILS cluster in Koha both admin as well as OPAC interface |
| Google Indic Transliteration | ILS cluster in Koha OPAC |
| Federated search system interface | Multilingual by using Discovery tools in VuFind |
| ISO 10646 UCS | Universal Character Set |
| ASCII | Multilingual standards for 8 bit code |
| ISCII | It covers 10 Indic languages derived out of Bramhi |

Table – 1 : Components of multilingual for libraries

Interoperability is a critical problem in the network environment especially when we are talking about the Digital Libraries with increase in number of diverse computer systems, software applications, file formats, information resources and users(Oakes & Xu, 2009). But it becomes more critical problem in Indian digital libraries, with having those much differences it has another sharing problem of resources from one language to another as resources at Indian libraries are present in many Indian languages viz. English, Hindi, Sanskrit, Marathi, Gujarati, Oriya, Bengali, Punjabi etc (Paolillo, Pimienta & Prado, 2007). Thus it has problem of interoperability between multilingual digital library resources. However there are so many true type fonts are being used to represent the Indian languages on web. But that's not sufficient tool to implement the multilingual (Peters, Braschler & Clough, 2012). ISCII is also being used as a standard to represent the Indian languages on the web as well on the database part. At the same time, users with other native languages

than that of the country under consideration may need more international languages, as for example, English Hindi or Bengali.

## 1.2 Development of Multilingual Environment for libraries

In general, the API of the middle layer should follow the Open - Closed principle, which states that software entities (modules) should be open for extensions, but closed to modifications. Being the system software, IM frameworks make extensive use of services provide by modern operating system (Shokouhi & Si, 2011). There are many languages are available in six cluster like integrated library system cluster, content management system cluster, college communication interaction cluster, federated search system cluster, learning content management system cluster and digital media archiving cluster (Mudawwar, 1997). These all cluster are managed through SCIM input method for solve the multilingual problem in college libraries under the university of Burdwan. SCIM input tools are easily managed the languages, fonts and script in table – 2 for developing the multilingual facilities both from staff-client as well as user interfaces and this table shows the 48 languages that can easily managed through scim tool.

| Sl. | Name of Languages | Sl. | Name of Languages | Sl. | Name of Languages |
|---|---|---|---|---|---|
| 1 | Amharic [41] | 21 | Hindi [13] | 41 | Tamil [33] |
| 2 | Arabic [42] | 22 | Japanese [14] | 42 | Telugu [34] |
| 3 | Armenian [43] | 23 | Kannada [15] | 43 | Thai [35] |
| 4 | Assamese[1] | 24 | Kazakh [16] | 44 | Tibetan [36] |
| 5 | Bengali [2] | 25 | Korean [17] | 45 | Uighur; Uyghur [37] |
| 6 | Burmese [44] | 26 | Lao [18] | 46 | Urdu [38] |
| 7 | Central Khmer [45] | 27 | Malayalam [19] | 47 | Vietnamese [39] |
| 8 | Chamic Languages [46] | 28 | Marathi [20] | 48 | Other [40] |
| 9 | Chinese [3] | 29 | Nepali [21] | | |
| 10 | Croatian [47] | 30 | Oriya [22] | | |
| 11 | Danish [48] | 31 | Panjabi; Punjabi [23] | | |
| 12 | Divehi;Dhivehi ; Maldivian [4] | 32 | Persian [24] | | |
| 13 | English [5] | 33 | Russian [25] | | |
| 14 | Esperanto [6] | 34 | Sanskrit [26] | | |
| 15 | French [7] | 35 | Serbian [27] | | |
| 16 | Georgian [8] | 36 | Sindhi [28] | | |
| 17 | Greek, Ancient (to 1453) [9] | 37 | Sinhala; Sinhalese [29] | | |
| 18 | Greek, Modern (1453-) [10] | 38 | Slovak [30] | | |
| 19 | Gujarati [11] | 39 | Swedish [31] | | |

| | | | | |
|---|---|---|---|---|
| 20 | Hebrew [12] | 40 | Tai Languages [32] | |

Table – 2: Multilingual languages represents through open source software

Different fonts and scripts are represents in the following way:-

1. Phonetic, inscript and itrans ; 2. Itrans, Unijay, Prabhat, Inscript, phonetic
3. Py, Pinyin, quick, tonepy, canjie and bopomofo; 4. Phonetic; 5. Ispell
6. q-sistemo, h-sistemo, h-fundamente, vi-sistemo, x-sistemo and plena
7. Azerty; 8. kbd; 9. Mizuochi; 10. Kbd; 11. Itrans, inscript and phonetic; 12. Kbd
13. Inscript, itrans, typewriter, phonetic and remington
14. Trycode, anthy and tcode; 15. Inscript, itrans and kgp
16. Kbd and Arabic; 17. Han2 and romaja; 18. Irt and kbd
19. Inscript, Mozhi, itrans and Swanalekha
20. Itrans, inscript and phonetic; 21. Rom and trad; 22. Itrans, phonetic and inscript
23. Jhelum, itrans, phonetic and inscript; 24. Isiri
25. Yawarty, phonetic, kbd and translit; 26. Harvard-kyoto; 27. Kbd; 28. Inscript;
29. Trans, samanala, wijesekhara-preedit, wijesekara, phonetic-dynamic, phonetic-static
30. Kbd ; 31. Post; 32. Sonla-kbd;
33. Typewriter, phonetic, itrans, lk-renganathan, inscript, tamil99
34. Rts, pothana, inscripts, itrans and apple; 35. Pattachote, tis820 and kesmanee
36. Ewts, tcrc and Wylie; 37. Kbd; 38. Phonetic
39. Tcvn, vni, han, nomvni, nomtelex, telex and viqr
40. Compose, latin-post, rfc1345, latex, latn-pre, syrc-phonetic and Unicode
41. Sera; 42. Kbd; 43. Kbd; 44. Kbd; 45. Yannis; 46. Kbd; 47. Kbd; 48. Post;

The updated SCIM Input Method provides efficient input facilities for the Bengali language in the Ubuntu operating system. This is the whole process of customizing and using the input software and is believed to be useful for anybody interested to develop a SCIM Input Method for their respective languages. All the languages are to be appeared in data entry interfaces for domain specific cluster and also see their fonts by seamless integration in Ubuntu operating system.

### 1.3 Methodology for Integration of Googleindictransliteration for libraries

Google indic transliteration is only available in online environment but this research work successfully integrated in ILS cluster Koha OPAC interfaces. Suitable and approapriate technological facilities are not available to the college users for their demands in different item types including books, journals, reference books and etc (Si & Callan, 2006). Library collections of different items not arranged in systematic order and not up-to-date OPAC. Require big room in library for large collections to the users. The college users can access the different documents from the existing catalogues (Ranta, 2004). Information mashup and cloud computing facilities is also available as mobile or android for the college libraries users and here display the cover images on online from amazon books and google books. The methodology is very simple to implement this tools in online public access catalogue for the libraries. Selection of standards and tools in multilingual transliteration from one language to another language and made a comparative study in two different aspects like comparative study of transliteration tools and ILS software in the domain specific cluster in different modules. The methodology in this fields are described in the following ways:

### 1.3.1 Selection of transliteration tools for libraries

Selection of transliteration tools in integrated library system cluster on the basis of global recommendations and local requirements for the college libraries. Localization refers to the process of adapting software to one specific language or culture. The locale model is one method to internationalize operating systems, and applications that run on it, and has been implemented on Unix (International Organization for Standardization, 1993). Only one locale can be specified for an application. Therefore, the user must explicitly switch the locale in order to use languages that are not defined in the current locale. This research work only select the matured level transliteration softwares these can be described as follows:

**Episimiotis**

Episimiotis is a tool for annotating a complex hierarchical and linguistic structure of any text and its user friendly. It was primarily designed for the tagging and analysis of errors made in written assessments by students of Modern Greek as foreign language by means of a predefined tagset. Linguistic annotation in texts is essential for the study of language and the development of NLP tools.

**Google Indic Transliteration**

It is one of the important approach in machine transliteration for managing the multiple languages from one language to another languages based on machine transliteration. The performance of machine translation and cross-language information retrieval depends extremely on accurate transliteration of named entities (Vijaya...[et.al], 2009).

**Multext Corpora**

Multext (Multilingual Text Tools and Corpora) is a recently initiated large-scale project funded under the Commission of European Communities Linguistic Research and Engineering Program, which is intended to address these problems.

**Semantex**

It is a version customized for triage on Arabic documents using entity identification, event extraction, and term translation. multilingual extraction allows non-linguists to conduct more precise, contextually accurate triage and information discovery. This helps ensure that scarce human language resources are used where most required.

**TransSMS**

The TransSMS service can be accessed via the Web or a Java enabled phone that has already downloaded the TransSMS client software. There is no difference in

terms of functionality between the two methods. Both include security features and text to speech translation capability. The user may request for the translated text to be sent as SMS to a recipient or request for a Call Back.

### 1.3.2 Comparison of tools for libraries

The comparison is made in two different aspects : (i) Comparison of transliteration tools (Table -3) and (ii) Comparison of integrate library software (Table 4). These multilingual tools are represents in the table – 3 for the selection of comprehensive transliteration tools on the basis of the global recommendations like IFLA Working Group and ILS-DI towards next level automated and digital library system and in such a way this research work calculate the score full supported tools considered as 1, partial supported tools considered as 0.5 and absence value represents as 0. In this way whose score is high this transliteration tool considered as most comprehensive for developing the transliteration in the college libraries under the university of Burdwan.

**(i) Comparison Results of transliteration tools :**

The results of transliterations tools in domain specific cluster prepared in tha table – 3. This research work select the matured level softwares for the college libraries are as follows:

| Sl. | Parameters | Episimiotis | | Google Indic Transliteration | | Multext Corpora | | Semantex | | TransSMS | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Support | Score | Support | Score | Support | Score | Support | Score | Support | Score |
| 1 | Peer-to-Peer (P2P) | Yes | 1 | Yes | 1 | Yes | 1 | Partial | 0.5 | Yes | 1 |
| 2 | Linguistic annotation | Partial | 0.5 | Partial | 0.5 | Yes | 1 | No | 0 | No | 0 |
| 3 | Text markup | No | 0 | Partial | 0.5 | No | 0 | No | 0 | No | 0 |
| 4 | Machine tanslation | No | 0 | Yes | 1 | No | 0 | Partial | 0.5 | No | 0 |
| 5 | Text encoding initiatives | Partial | 0.5 | Yes | 1 | Partial | 0.5 | No | 0 | Partial | 0.5 |
| 6 | text analysis | Yes | 1 | Yes | 1 | No | 0 | Partial | 0.5 | No | 0 |
| 7 | Multipurpose Internet Mail Extensions | No | 0 | Partial | 0.5 | Partial | 0.5 | No | 0 | Yes | 1 |
| 8 | User interface | No | 0 | Yes | 1 | No | 0 | Partial | 0.5 | No | 0 |
| 9 | Universal Character Set or UTF | Partial | 0.5 | Yes | 1 | Partial | 0.5 | No | 0 | No | 0 |
| 10 | Localization | Partial | 0.5 | Yes | 1 | No | 0 | No | 0 | Partial | 0.5 |
| Total Score (out of 10) | | Episimiotis Score : 4 | | Google Indic Transliteration Score : 8.5 | | Multext Corpora Score : 3.5 | | Semantex Score : 2 | | TransSMS Score : 3 | |

Table – 3: Comparison results of transliteration tools for CLBU

From the above table -3 it can shows that the score of transliteration tools like Episimiotis Score : 4 out of 10, Google Indic Transliteration Score : 8.5 out of 10, Multext Corpora Score : 3.5 out of 10, Semantex Score : 2 out of 10 and TransSMS Score : 3 out of 10. So, the highest score is Google Indic Transliteration tools as compared to other transliteration tools in the above table. Obviously, it can conclude that Google Indic Transliteration is the most comprehensive machine transliteration tools for designing and developing the college libraries under the university of Burdwan because it can possible to integrate the multilingual transliteration in ILS OPAC like Koha.

**(ii) Comparison Results of ILS Softwares**

Comparative study is prepared of six open source matured ILS software for the selection of most comprehensive software to managed the transliteration in Koha OPAC and in this respect parameter is selected on the basis of global recommendations like ILS-DI and IFLA Working Group recommendation. Most comprehensive parameters are API code, CSS, Java script, Unicode, perl, masthead, system preference, change languages, transliteration and search box and these parameters represents in the table – 4. Here 0 represents absence value, 0.5 represents partial value and 1 represents presence value.

| Sl. | Parameter | Score of open source software against in multilingual | | | | | |
|---|---|---|---|---|---|---|---|
| | | Emilda | Evergreen | Koha | NewGenLib | OPALS | WEBLIS |
| 1 | API code | 1 | 0.5 | 1 | 1 | 0 | 0 |
| 2 | CSS | 0.5 | 0 | 1 | 0 | 0.5 | 0.5 |
| 3 | Java script | 0 | 0.5 | 0.5 | 1 | 1 | 0.5 |
| 4 | Unicode | 1 | 0.5 | 1 | 1 | 0 | 0.5 |
| 5 | Perl | 0.5 | 0 | 1 | 0 | 0 | 0 |
| 6 | OPAC customization scopes | 0 | 0 | 1 | 0.5 | 0 | 0 |
| 7 | System administration | 1 | 1 | 1 | 1 | 0 | 1 |
| 8 | Change languages | 0.5 | 0.5 | 1 | 1 | 0.5 | 0.5 |
| 9 | Transliteration | 0 | 0 | 1 | 0 | 0 | 0 |
| 10 | Search box | 0 | 1 | 1 | 1 | 0 | 1 |
| Total Score (out of 10) | | 4.5 | 4 | 9.5 | 6.5 | 2 | 4 |

Table – 4: Comparison results of ILS Softwares for CLBU

From the above table it can shows that the Koha gives highest score 9.5 out of 10 whereas NewGenLib 6.5 out of 10; Emilda score 4.5 out of 10 ; WEBLIS score 4 out of 10 ; Evergreen score 4 out of 10 and OPALS score 2 out of 10. So, obviously it can indicates that transliteration is easily possible in Koha OPAC interface for

designing and developing the college libraries under the university of Burdwan.

### 1.3.3 Integration Methods in Koha

The integration method of Google Indic Transliteration in Koha OPAC is to make in a simple way. In this section configure OPAC related seven files namely koha-tmpl/opac-tmpl/prog/en/css/opac.css,/opac-tmpl/prog/en/includes/doc-head-close.inc,koha-tmpl/opac-tmpl/prog/en/includes/masthead.inc,/prog/en/js/googleindictransliteration.js,opac/opac-main.pl,opac/opac-search.pl,koha-tmpl/opac-tmpl/prog/en/js/googleindictransliteration.js. After that create a new system preferences related with Google Indic Transliteration if on it the transliteration will appear in Koha OPAC pages as masthead.

### 1.3.4 Development of Googleindictransliteration in Koha

The google transliteration gives one java file and configured this file according to the languages code which is essential in koha OPAC for machine transliteration from one language to another languages. Configure the java file under the /usr/share/koha/opac/htdocs/opac-tmpl/prog/en/js/ googleindictransliteration.js the following java file will generate in the figure – 1.

```
      // Load the Google Transliteration API
      google.load("elements", "1", {
            packages: "transliteration"
          });

        function onLoad() {
        var options = {
          sourceLanguage: 'en',
          destinationLanguage: ['am',
'ar','bn','el','fa','gu','hi','kn','ml','mr','ne','or',
'pa','ru','sa','si','sr','ta','te','ti',
'ur','zh'],
          shortcutKey: 'ctrl+g',
          transliterationEnabled: true
        };

        // Create an instance on TransliterationControl with the required
        // options.
        var control =
            new google.elements.transliteration.TransliterationControl(options);

        // Enable transliteration in the textfields with the given ids.
        var ids = [ "transl1" ];
        control.makeTransliteratable(ids);

        // Show the transliteration control which can be used to toggle between
        // English and Hindi and also choose other destination language.
        control.showControl('translControl');
      }
      google.setOnLoadCallback(onLoad);
```

Figure – 1 : Google Transliteration Java File

Figure – 1 : Google Transliteration Java File

Write here 22 languages and default languages is English. The name of twenty languages are Amharic: 'am', Arabic: 'ar', Bengali: 'bn', Chinese: 'zh', Greek: 'el', Gujarati: 'gu', HIndi: 'hi', Kannada: 'kn', Malayalam: 'ml', Marathi: 'mr', Nepali: 'ne',

Oriya: 'or', Persian: 'fa', Punjabi: 'pa', Russian: 'ru', Sanskrit: 'sa', Sinhalese: 'si', Serbian: 'sr', Tamil: 'ta', Telugu: 'te', Tigrinya: 'ti', and Urdu: 'ur'. Short key of this tools is ctrl+g for google indic transliteration.

### 1.3.5 Testing for libraries

This research work to add the Google Indic Transliteration tool to the masthead on the OPAC. Google indic transliteration is web 2.0 features in integrated library system cluster. This tool transliterates text in the source language to a destination language selected from a drop-down list. The transliterated expression can be then be used as a search expression. In this respect the figure – 2 will generate as follows :

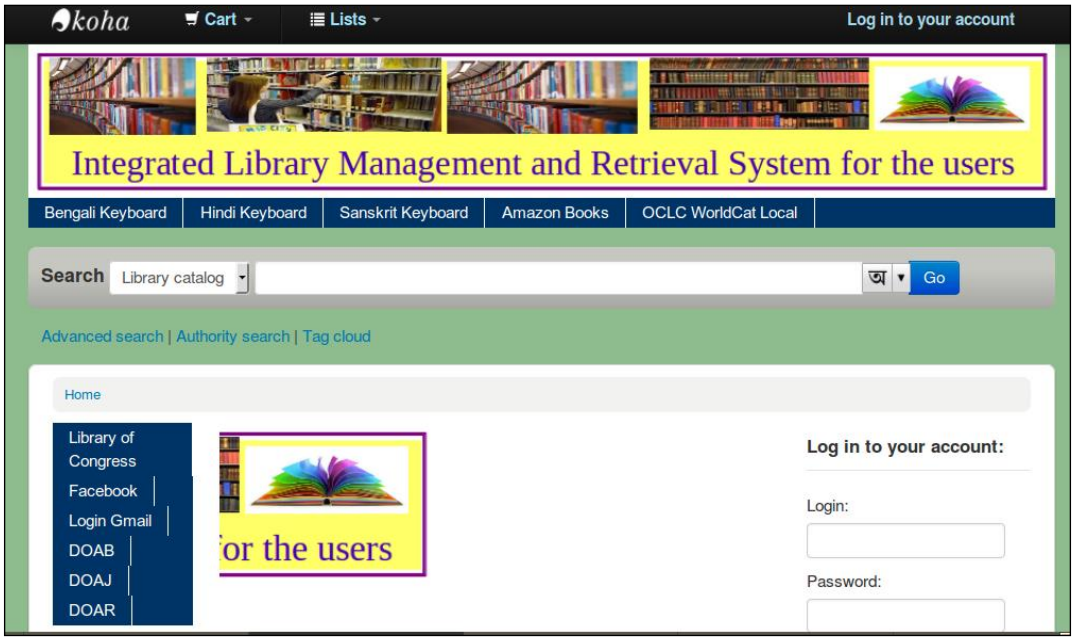


Figure – 2 : Testing of Translation in Koha OPAC

### 1.3.6 Results for libraries

Users and librarians can managed the twenty two languages from the library OPAC in Koha including Amharic, Arabic, Bengali, Persian, Greek, Gujarati, Hebrew, Hindi, Kannada, Malayalam, Marathi, Nepali, Oriya, Punjabi, Russian, Sanskrit, Serbian, Sinhala, Tamil, Telugu, Tigrinya and Urdu (Yuwono & Lee, 1997). All the languages are to be access through Google input tool and this developing made on Ubuntu operating system due to its higher security rather than Windows operating system. But this google input tool support both the operating system yet this research work select only Ubuntu operating system. Koha is fully support the Unicode based standards for manage the multilingual resources and all the language code available in online environment to access from the library OPAC pinpointedly, exhaustively and expeditously. But here internet connection is mandatory for translating the resources from source language to destination languages. The figure – 3 will represents the transliteration is possible from one language to another languages and it can convert in type word languages (Viles & French, 1995). If ignore the transliteration from the

Koha library OPAC press the ctrl+g and again type in English for search the documents to retrieved it from the specific library. Here English is default languages because this file known as java base googleindictransliteration file. After testing the Google Indic Transliteration in Koha OPAC the all language will appear and translate it from English to Bengali and also other 22 languages. This is the most easy process to integrate in Koha OPAC (Figure – 3) for managed the multilingual transliteration. The results of translate from English to Bengali and to ignore the transliterate press ctrl+g. This is the most innovative features towards next level automated and digital library system.

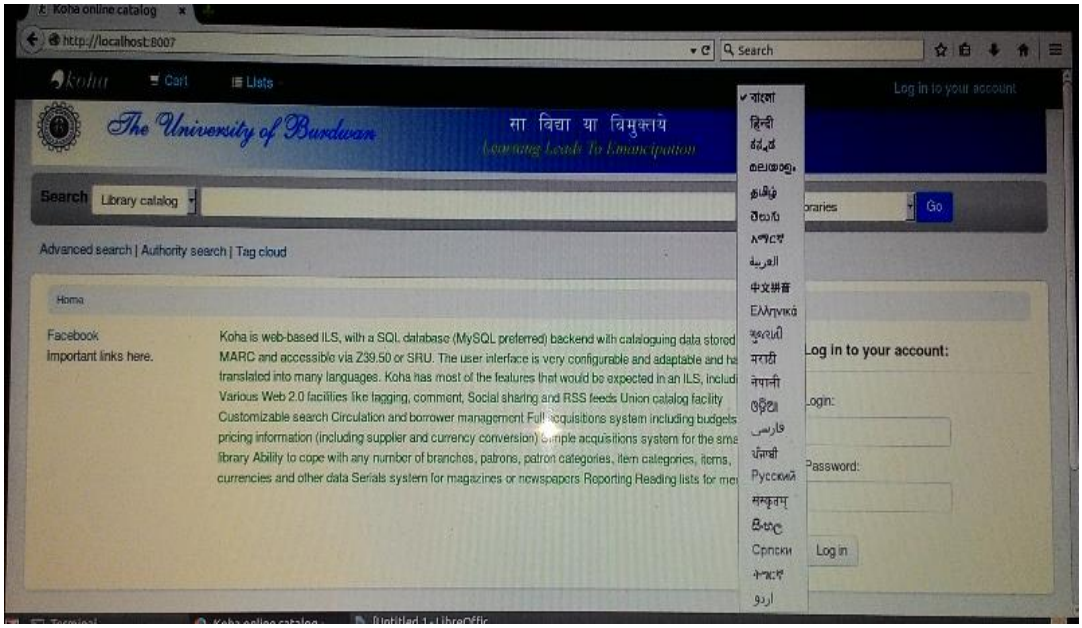


Figure -3: Google indic transliteration in Koha OPAC

The most of the college libraries are facing problem in Bengali transliteration but this research work try to solve this problem through Google Indic Transliteration tool in Koha OPAC interface. The transliteration model also performed better when compared to Google Indic transliteration. But the fact is that the Google system is designed for general transliteration whereas the model presented here is trained exclusively for Indian names and places. It is concluded that this transliteration model is applicable for the languages which have the same alpha-phonetic sequence in both source and target languages. This transliteration framework is designed on the basis of global recommendations for designing and developing the college libraries under the University of Burdwan.

### 1.4 Development of Multilingual Framework

Multilingual framework is required in domain specific cluster. Many books are available in the college library in different languages such as Bengali, Hindi and Sanskrit and etc but how to managed these types of books in the college library. It is possible to developed the multilingual framework by using the open source software Koha. In most of the college libraries are to be required Bengali languages because

there is no standard software in the college environment. This research work development the multilingual framework in the following procedures :

### 1.4.1 Bengali Language Installation in Koha

Bengali language installation in Koha both for the OPAC and Intranet user interfaces at any time to a running koha installation from the directory /usr/share/koha/misc/translator. First configure and using the two commands by terminal to specify the location of Koha perl modules and of the koha-conf-site.xml.in file and open Applications > Accessories > Terminal and use the following commands :

```
sudo su
export KOHA_CONF=/etc/koha/sites/library/koha-conf.xml
export PERL5LIB=/usr/share/koha/lib
cd /usr/share/koha/misc/translator
perl translate install bn-IN
```

### 1.4.2 Settings Multilingual System Preferences in Koha

Global system preference settings for Bengali language only on the Bengali options under the l18N/L10N both for Koha admin as well as OPAC interface. The Figure – 4 indicate the system preference options in integrated library system cluster.

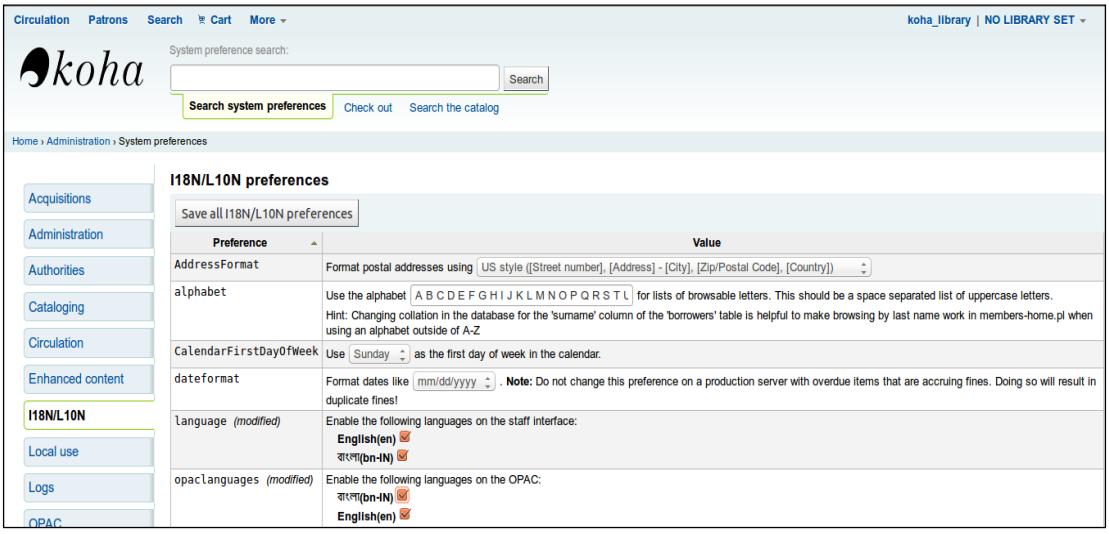


Figure – 4 : Setting system preference in Koha for Bengali language

### 1.4.3 Translate in Koha Modules

Configure the Koha in Bengali language under the directory of /usr/share/koha/intranet/ htdocs/intranet-tmpl/prog/bn-IN/modules and manually translate the each file in Koha admin interface. Also translate the OPAC interface in Koha under the directory of /usr/share/koha/opac.

### 1.4.4 Bengali Interface in Koha

The Figure – 5 reveals that the Bengali interface in Koha administration and this will appear after translate the all modules files effectively and efficiently in the integrated library system cluster. This interface is helpful only for the college librarians but not the users. It also affect the library professionals those are interested in open source software.


Figure – 5 : Bengali interface in Koha Admin Interface

## 1.5 Bengali Data Entry Process in Koha

Data entry for bibliographic descriptions in the MARC 21 format is possible in two ways Avro-phonetics and Virtual Keyboards. The facility of customization truly characterizes open source software. Koha has tremendous possibility in automating College libraries in India. This section deals with the customization of Kohlrabi for use in College libraries in West Bengal. In West Bengal, most of the College libraries require facility to process, store and retrieve Bengali script based documents. Apart from this necessity college libraries require Bengali Script based user interface and need export and import facility of Bengali script based documents in ISO-2709 format. Keeping in view all these facts, a project on customizing Koha has taken by the author to support the above mentioned requirements of College libraries in West Bengal. The first problem encountered in this endeavor is that the Koha is not Unicode- compliant. Although all the software required to run Koha (Apache, MySQL, PERL) allows universal character set, Koha itself is not Unicode compliant and therefore Koha source code requires to be modified to allow processing of Bengali script based information objects (Ruiz & Chin, 2010). This problem is solved through the development of a Unicode-compliant and Bengali script based theme for Koha. This theme can be installed separately over the top of regular Koha installation. Administrator of Library automation system (or Koha) can configure Koha easily to use this theme. Change of this theme to the default theme of Koha is the matter of a click. It means any time administrator can roll back to the default theme of Koha. The

data entry is also possible by using avro phonetic keyboard on ubuntu interface. In this way the Koha – 3.X is support the ubuntu linux operating system so it can easily entered the data in Bengali through avro phonetic and it can visible in staff client and opac interface. Simultaneously it can also managed the multilingual resources and also their fonts. Now, SCIM input method is an important tools in Ubuntu operating system which can easily managed the Bengali Script in College libraries under the University of Burdwan for designing the integrated library management system and retrieval system.

### 1.5.1 Data Entry through Ibus Avro Phonetics

The data entry is also possible by using avro phonetic keyboard on Ubuntu interface. In this way the Koha is support the Ubuntu Linux operating system so it can easily entered the data in Bengali through avro phonetic (See Figure -6) and it can visible in staff client and OPAC interface. Simultaneously it can also managed the multilingual resources and also their fonts. Now, SCIM input method is an important tools in Ubuntu operating system which can easily managed the Bengali Script in College libraries under the University of Burdwan for designing the integrated library management system and retrieval system.

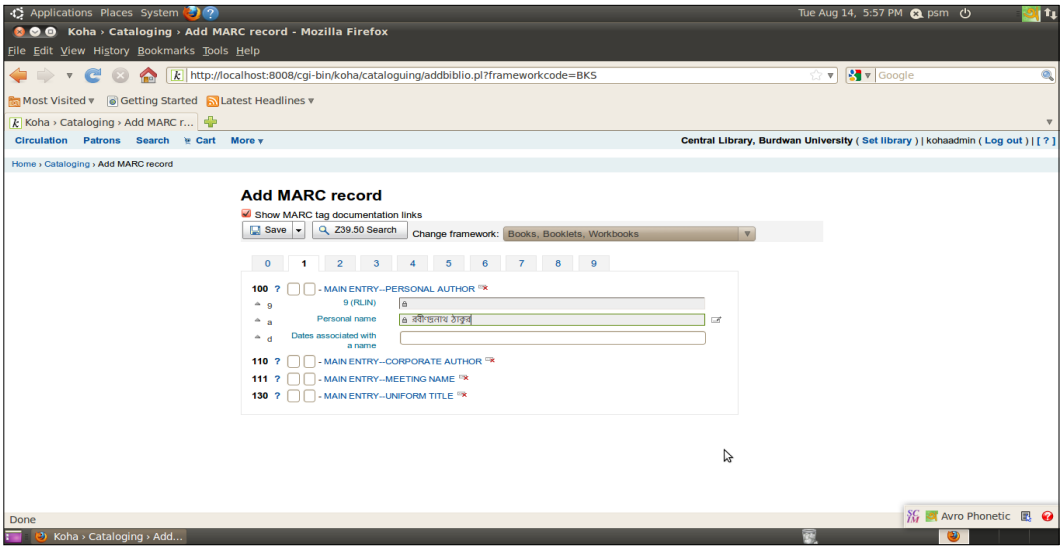


Figure -6 : Avro Phonetics in Koha on Ubuntu

### 1.5.2 Data Entry through Virtual Keyboard

In this section only highlights the Bengali data entry framework. Data entry is also possible through virtual keyboard in domain specific cluster. Integrated library system cluster consists of two interfaces such as koha admin and Koha OPAC interface in the college libraries because virtual keyboard is easily managed the bengali script and language both for librarian and OPAC interface in Koha and not only support in integrated library system, it also support the other five domain specific cluster (Roberson & Walker, 1994). Virtual keyboard can be use in two ways like click on mouse and type from the computer keyboard (Rountree, 2012). Spelling

correction is also possible in each words because its appear nearest spelling and here select the correct spelling during typing. Obviously, it can save the time of the librarians and college users. Integrated of Virtual keyboard in Koha OPAC only by clicking on mouse. Regional language searching searching is one of the important problem of every library, so this research work solved this problem by configuration of Zebra indexing in Koha in the following ways:

## I. Regional Language Searching in Koha

Library is the only place where users both students and teachers also access and searching the library materials in their own language. This research work is successfully searching all the documents in Koha by Zebra indexing. Users and librarians of all the colleges can be easily search in different item types of different languages which enter in Koha both for bibliographic and authority data. Configuration of Zebra for searching the regional languages in Koha both for librarian and OPAC interfaces. In this stage first open the zebra database in Koha through terminal use the following command :

```
sudo su
gedit /etc/koha/zebradb/etc/default.idx
```

Now, here to find out the important line "**charmap word-phrase-utf.chr**" and inserting by # symbol which represnts in the following line:

```
# Zebra indexes as referred to from the *.abs-files.

#  $Id: default.idx,v 1.10.2.1 2004/09/16 14:07:50 adam Exp $

#


# Traditional word index

# Used if completenss is 'incomplete field' (@attr 6=1) and

# structure is word/phrase/word-list/free-form-text/document-text

index w

completeness 0

position 1

alwaysmatches 1

firstinfield 1

#charmap word-phrase-utf.chr

icuchain words-icu.xml
 [ add the following line ]
```

```
# Phrase index

# Used if completeness is 'complete {sub}field' (@attr 6=2, @attr 6=1)

# and structure is word/phrase/word-list/free-form-text/document-text

index p

completeness 1

firstinfield 1

#charmap word-phrase-utf.chr

icuchain words-icu.xml
 [ add the following line ]

# URX (URL) index

# Used if structure=urx (@attr 4=104)

index u

completeness 0

charmap urx.chr
```

Finally, start the Zebra indexing in Koha from the terminal by using the following command:

```
sudo koha-rebuild-zebra -v -f library
```

All the regional languages are searching by Koha for the students in college libraries. This can be done through the Zebra indexing due to Koha is fully support the Zebra. In most of the college libraries are easily manage the Bengali language. So, obviously, it can searching and browsing the different items which available in the academic libraries.

## II. Search Results of Regional Language

College libraries can easily search the regional languages of books and other library materials. The number of books are count in a single window of different wise and branch wise also. Regional language setup is start from the Koha administration under global system preferences. Search results display in different sets and in different formats such as normal view, ISBD and MARC view. Each an every records is easily searching both the librarian as well as OPAC interfaces. The search results are described in the next chapter of features of the integrated framework due to all the important results with access point discussed in this section. Now, the Figure – 7 is represents the search results of regional language and here regional language is Bengali because here most of the people speak in Bengali language. This framework is more helpful to all the libraries.

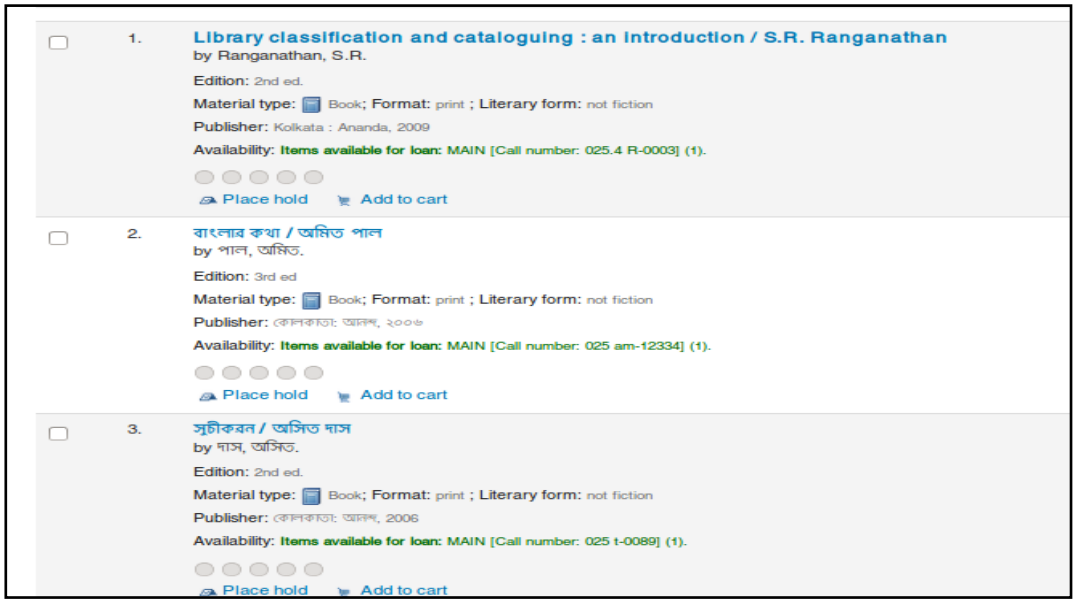


Figure – 7 : Search results of regional languages in Koha for libraries

### 1.6 Development of Multilingual Digital Resource Management

Development of multilingual in digital media archiving cluster basically in two areas like metadata entry in DSpace by Bengali language. On the otherhand metadata entry in Greenstone by Bengali language is not possible but Bengali language support in user interface (Stiller, Gade & Petras, 2013). Greenstone support lucene and MGPP indexing tools and DSpace only support the lucene indexing tools (Powell & Fox, 1998). Both DSpace and Greenstone multilingual full text digital resource can be managed through search browsing and browsing classifiers. College library can easily managed the digital resources by using these two open source software. There are three interface in Greenstone such as librarian interface, greenstone editor for metadata schema and greenstone user interface. Apart from these DSpace consists of three interface including DSpace admin, DSpace user and DSpace XMLUI based interface in developing the digital media archiving cluster for the libraries.

#### 1.6.1 Metadata Entry in Bengali of DSpace

Metadata means data about data. DSpace support the multilingual in Unicode based open source software. College libraries are facing the problem the management of Bengali language full text resources and this can solved by using the DSpace in metadata. There are three types of metadata can be managed in digital library environment including administrative, structural and qualified dublin core metadata (Ponte & Croft, 1998). The all the metadata is easily managed in Bengali language and other languages because its unicode based supporting software. Designing of user interface in DSpace is very easy because its support the html format and here just write the html code. It helps to preserve the digital documents in college libraries and search, browsing and indexing both alphabetical in descending and ascending order.

To easily find the creators, title and subjects in different metadata schemas because its support the dublin core metadata schema. Databse backup and restorations of metadata is also possible through postgresql database management system. Multilingual data entry is to be made through the different languages on Ubuntu just on the bengali language font both in mouse of computer and keyboard comfortable. It is managed both structural and descriptive metadata in digital library system. Users can access the bengali documents from the DSpace repositories in different ways including browsing, searching, indexing and download the full text documents. Indexing is very approapriate in searching because its support the lucene indexing tool both for the users and DSpace admin interfaces. Change the language from source to destination from the XMLUI interface, here change all the message keys in different files, directories and sub-directories because its support qualified Dublin core metadata schema. Crosswalked and interoperability is also possible from the different system during the data conversion.

### 1.6.2 Multilingual Search Results for libraries

Users can search the document in multilingual data format and they get their necessary search documents easily because here automatically indexing system tools are to be used (Buttenfield, 1999). Also users search the different languages such as Bengali, Hindi and Sanskrit and other languages. The search results of DSpace user interface in Bengali are retrieved in the Figure – 8 to choice their full text documents as well as metadata related on a particular college resources in digital media archiving areas (Fuhr, 2007). Only display the results in user interfaces but not edit or delete the documents or item from the databases. But in case of admin interface of DSpace or Greenstone the search, edit and delete is possible but here required suitable login and password.

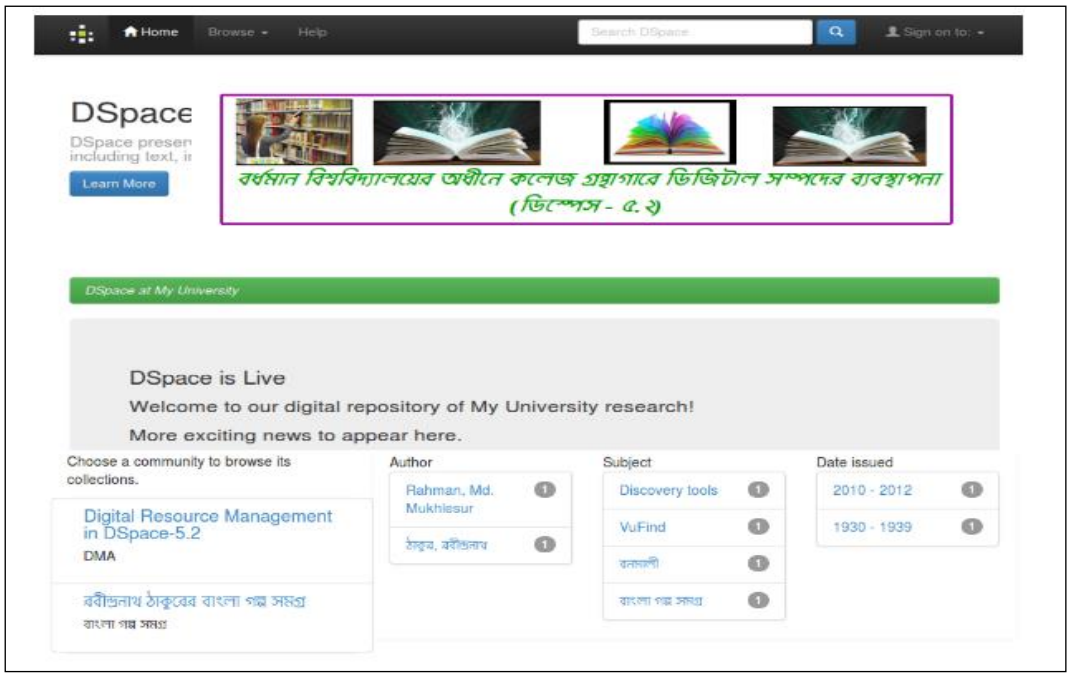

Figure – 8 : DSpace multilingual data for CLBU

Greenstone support the multilingual in digital media archiving cluster and this will represent in the Figure – 9 to managing the digital resources in the college libraries and also managed the full text Bengali, Hindi, Sanskrit and etc. for the Greenstone librarian interface.

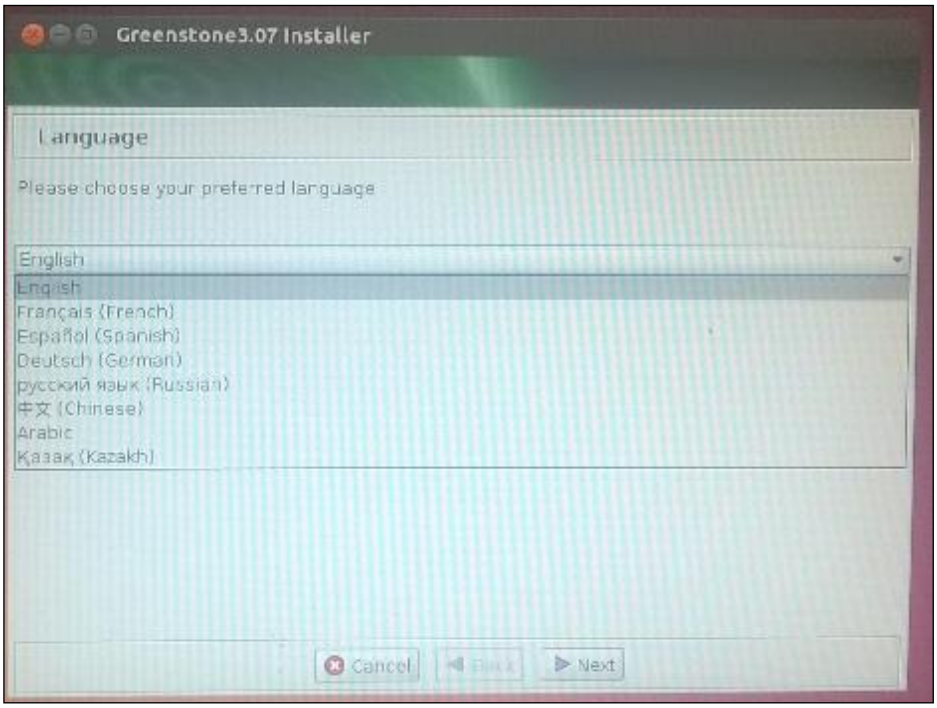


Figure – 9 : Greenstone Multilingual windows for CLBU

Multilingual is Greenstone user interface represents in the Figure – 10 in digital resource management for the college libraries in different types of item types such as Books, Journals, conference proceedings and etc.

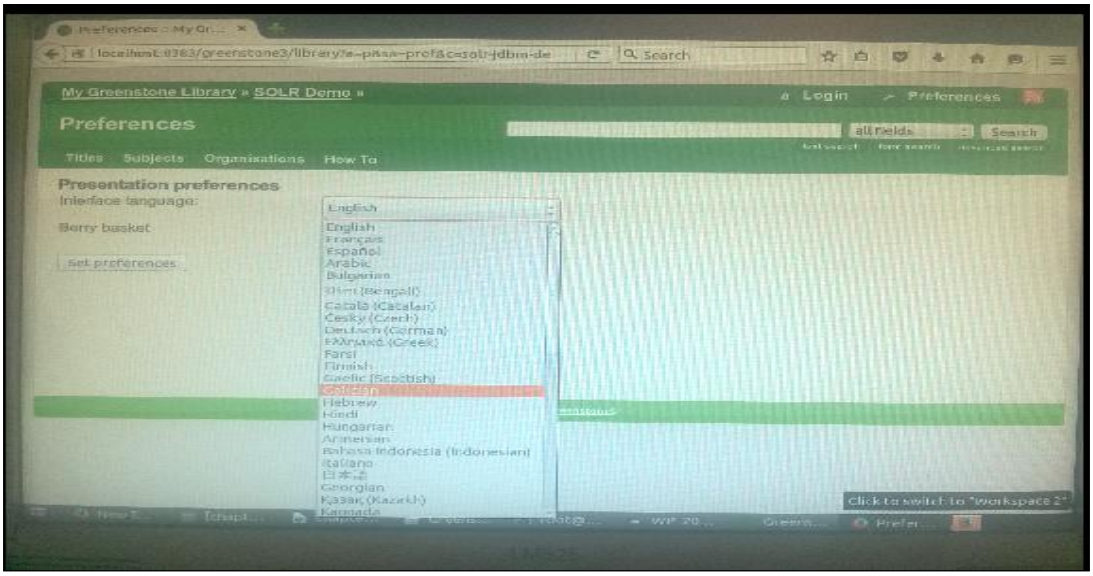

Figure – 10 : Multilingual interface in Greenstone user

Multilingual is also support in Greenstone user interface in digital media archiving cluster and Bengali Language represents in the Figure – 11 for the college users as well as library professionals. Greenstone is the most popular software in the digital library environment because here possible to create new indexing and browsing classifier both admin and user interface.

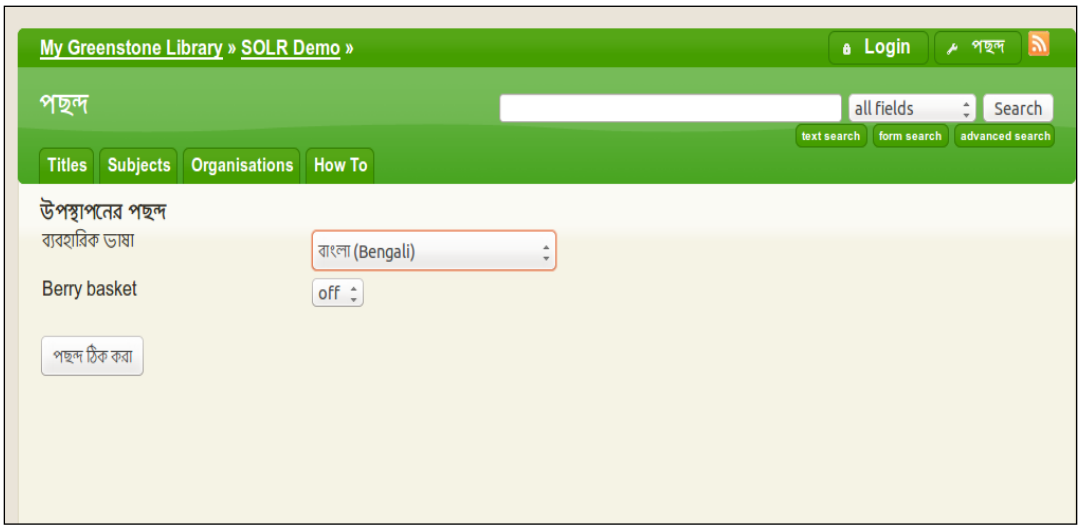


Figure – 11 : Bengali interface in Greenstone

Hindi language is also managed in Greenstone (Figure – 12) and users can access their necessary documents. There are different types of search facilities in user interface including advanced search, phrase search, stem searching, boolean searching and etc. for the college libraries affiliated to the University of Burdwan.

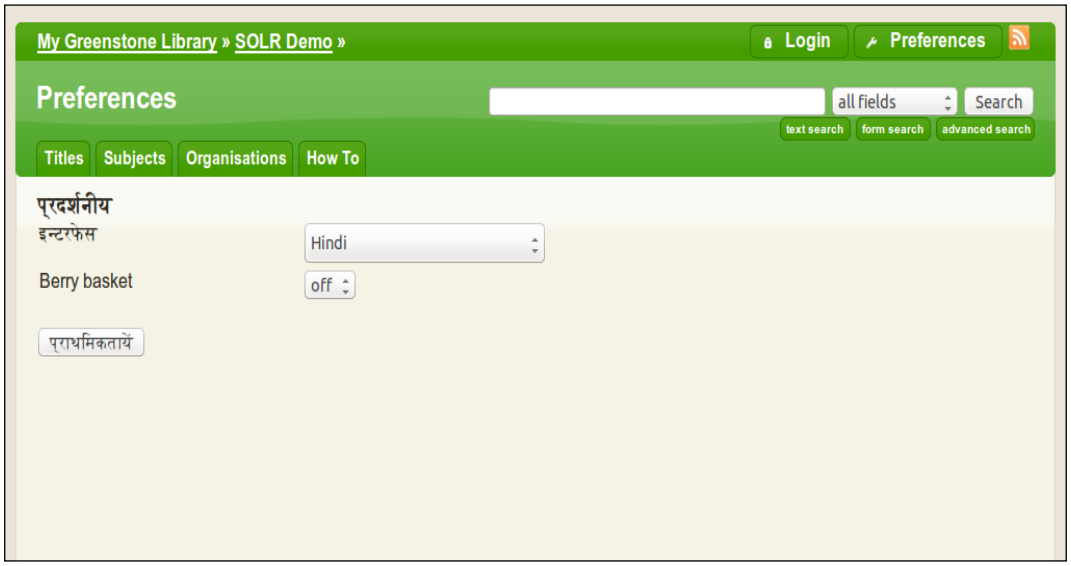


Figure – 12: Hindi interface in Greenstone

## 1.7 Management of Multilingual for libraries

There are many languages are available in the multilingual environment and

these languages can be managed through open source softwares. Apart from these discovery tool is also the important for the management and retrieved the full text as well as bibliographic information in the domain specific cluster. The management and development of multilingual for the different aspects in the libraries are described as below:

### 1.7.1 Development of Multilingual in Federated Searching for libraries

Federated search system development is also an important task in college libraries for grouping the collections, access the collections and download the collections and its retrieve the relevant results. Multilingual development is also possible through federated search system (Cox, 2007). Therefore need to address three major issues: how to represent the collections, how to select suitable collections for searching; and how to merge the results returned from collections (Rounter, 2012). Federated search system helps to college users they can access their necessary documents through information retrieval technology and it allows the search of different types of digital resources and full text documents which available in directory of open access repositories (Gazen & Minton, 2005). Aggregates the search results from the particular repositories and access the documents for the users one query, here retrieve all the relevant information that harvest from the other institutional repositories (Shokouhi & Si, 2011). Bibligraphic data access through the Web-enabled architecture in integrated library system using the Z39.50 server and SRU/SRW. It also manage the web-based search engines like Google, Yahoo-pipe and Rollyo to improve the relevance and accuracy of different search terms and its reduce the time for the users (Tran, 2011). Retrieve only the relevant information to the researchers and users from the multiple databases available in online environment. Google custom search engine is also support the federated searching because its retrive only those information which integrated the custom search engine in college library of different areas automatic indexing, customization, theme change, widget facilities, tinyurl and etc for specific types of resources. Mulitilingual resources are to be managed by using the federated searching tools like VuFind. Also multilingual searching is possible through OAI-PMH related harvesters like open conference system, open journal system, open harvester system and open monograph press. Federated search system is also possible through Z39.50 server in the domain specific cluster (Rogati & Yang, 2003). Multilingual data import from the other library OPAC by the Z39.50 server for developing the federated search system in the college libraries. Open monograph press manage the Books or monograph because its web-enabled architecture on Ubuntu operating system. The main purpose of this tool is to create the website with catalogue in different item types including catalog of books, distribution and handle the edited multi-volumes with different authors for each an every chapter of books. It also involve the bibligraphic description including editors, authors, indexers, book publication and reviewer.

Traditional search engines are not support the mulitilingual interfaces due to lack of technical knowledge in web visible content. Resource discovery interface automatically indexing the document through the algorithm system in the areas of library thing (Craswell & Hawking, 2000). Efficient results retrieve from the modern search engines by using the application programming interface and retrive the correct

items for one single search (Gazen and Minton, 2005). Date range and advanced search facilities is also available for the search terms. Information mashup and cloud computing can managed by the wrapper and resource discovery tools in different subject areas from the hidden information sources (Liu et al., 2001). The bibliographic and authority information can be forwarded to another person from the mail server after that the client user download and access the documents (Voorhees et al., 1995). Natural solution is to be made from the ranked lists of retrieving results in a particular repositories. Web content is visible through the discovery tools and it managed multilingual (Baeze-Yates & Ribeiro-Neto, 1999). Recent updated multilingual resource is manage by really simple syndication which represents the virtual big document in semanti web (Yuwono & Lee, 1997).

#### 1.7.1.1 VuFind Multilingual Discovery tools for libraries

"VuFind Rocks the House" by Roy Tenant. Multilingual document can be managed by using the VuFind discovery tolls. Also users can access the documents which are available in the databases. Now, libraries are turning into access point libraries from big warehouse type of libraries. Retrieved of multilingua electronic resources are rapidly developing and changing in the discovery layer services (Mizera-Pietraszko & Zgrzywa, 2010). To meet up the ever increasing demand of digital resources, libraries throughout the world are expanding their horizon in subscribing digital resources for their clients (Osborne & American Library Association, 2004). At the same time managing and providing access to those digital resources is also a major concern for the library and informational professionals worldwide (Powell & Fox, 1998). With the development of web environment, knowledge management in libraries became convenient both for professionals and the users. The multilingual interface of VuFind discovery tools presents in the Figure -13. This tool is considered as resource discovery because not only benefited the students but also helpful for the researchers. This can easily managed the citation styles in multilingual document for the different subject areas.

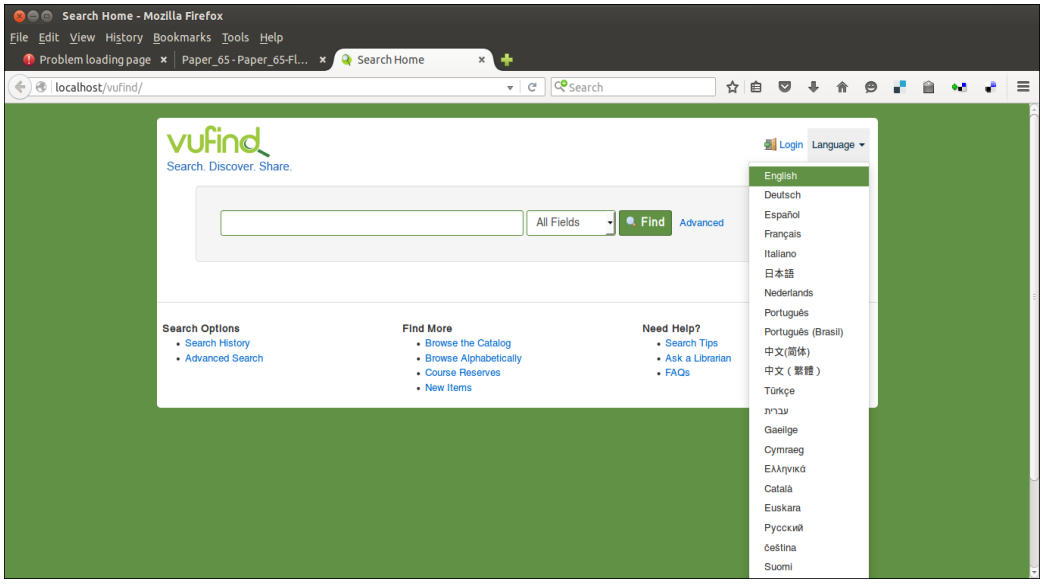

Figure – 13: Multilingual in VuFind discovery tool

#### 1.7.1.2 Multilingual Retrieval in OAI-PMH tools for libraries

OAI-PMH stands for open archive innitiative for protocol metadata harvesting. It supports the Unicode based mulitilingual standards for managing the federated resouces (Yang, 1999). Metasearching is also known as federated searching. Apart from this federated search system known as other name including cross searching, broadcast searching and other name. It is the powerfull search in efficientcy from the multiple web information resources. Advanced users can access the resources and some new users also upload and download their bibligraphic and authority information (Robertson & Walker, 1994). It is fully manage the multilingual resources due to its support the Unicode based standards. Users can easily harvest the multilingual resources from the institutional repositories which available in online digital resources. Z39.50 server is also an another federated multilingual searching tool in integrated library system. Integration of Z39.50 in Koha librarian interface is mandatory during web installation and search the bibligraphic documents as title, author, ISBN and etc for import the information from the other library OPAC. Obviously, it can save the time of the library professionals to manage the library resources. There are four tools are available in the Website of Public Knowledge Protocol including open harvester system, open conference system, open journal system and open monograph press but this research work selected only the open harvester system. On the otherhand only discussed the application of other three OAI-PMH related tools and open conference system to generate the website related with conference that allows searching both simple and advanced by using the fields of crosswalked related harvested archives. The federated search system is retrieved the relevant information in multilingual format (Ponte & Croft, 1998). Multilingual searching is also possible and retrieve the right information which wants to the users in a library.

Several measures, such as precision, recall, term overlap, and efficiency have been used to evaluate searching in bibligraphic databases (Viles & French, 1995). When applied to searches for specific facts in a full-text database, these measures the approapriate. Most commercial text retrieval systems use files to improve retrieval speed (Turtle, 1990). Full-text information retrieval systems have always attracted special attention due to the complexities involved in the storage, processing, and retrieval of large volumes of information. Full-text searching is likely to become an even more important activity in the future as the amount of information (Savoy & Rasolofo, 2000). The technology that makes its possible is the client-server model of networking, which essentially separates the user interface from the database and its suitable software. The client server approach allows the interface to reside on the local machine, rather than to be downloaded from the host, and requires a communication or protocol to interact with the search engine. Several organizations have developed specialized user interfaces for the Internet (Yuwono & Lee, 1997). The notion of interoperability between different database systems is to attractive that it has generated many different attempts to achieve multilingual standards (Zhu, 2005).. This aim mainly to perform two sets of functions first to enable machines and second information systems to be able to communicate with one another, to share and exchange data and so on (Zhai & Lafferty, 2001). It is also enable users to have access

to more than one information system using the harvester techniques and the OAI-PMH base URL.

#### 1.7.1.3 Multilingual Data Import through Z39.50 Server

The purpose of this section is importing and editing the bibliographic as well as authority based mulitilingual data for search and retrieval of records in the database (ANSI/NISO, 1995). The library today is being revolutionized with advancement of information technology and new tools and techniques. The future librarian may be designated as cybrarian or cyber librarian, as librarian has to provide information service from a large number of documents which are published in digital form and available in Internet. Now a days significant number of documents are now available in the Internet as free of cost. So, the college librarians may find some benefits if a computer system provided to the library in the areas of domain specific cluster. So a library may think to reorient its activities with the help of modern technologies. It may not be far away when a large number of students will demand computerized service from a college library. Bibligraphic records and authority records import from the Z39.50 client server architecture because this architecture web-enabled and here users can access the online information by using the Z39.50 server (Bergman, 2001). Koha is fully support the Z39.50 server and it also support the MARC 21 records as OXX-8XX fields except in 9XX because its consider as local resevation of a particular library. Information organization and retrieval is possible in the level of interoperability and crosswalked for college libraries (Buchinski, Newman & Dunn, 1976). Data access from different web server including library of congress is the world largest collection of items like Books, monographs, maps in multiple subject fields. Also, Koha is giving the many facilities that one can migrate from an existing ILS system to Koha and it also has the infrastructure to develop a digital library. All the MARC 21 tags can be shown through the structure parameter. Here one can ignore the tags which does not match to there requirements and edit the subfields of the required tags. As for example,this research work can take the MARC tag 245 for title statement. It can be mapped as below and here tab denotes the place where staff-client want to keep these information and -6 means hidden the subfield. One thing is essential, after mapping relationship between MARC and Koha field one should check shether it is correctly mapped or not in the MARC check parameters. Users can search the catalogue, request for items, also can know the details of books issued to them, membership details through this interface, locally as well as through Internet. Ranging from the name, address and designation, such details of the users to the items issued to them can be known. Acquisition process is also performed by Koha through using the Z39.50 server in multilingual resources. Koha provides an option for the database of vendors, through which one can place order for items to them. When data import from the other library OPAC through Z39.50 server, all the tags, fields, subfields and their related tabs are to be imported into Koha for copy cataloguing that can be represents as follows:

| Tag-subfield | Koha field | Tab |
|---|---|---|
| 6 | | 2 (-6) |
| 8 | | 2 (-6) |
| a | bibliotitle | 2 (0) |
| b | biblio.subtitle | 2 (0) |
| c | | 2 (0) |
| f | | 2 (-6) |
| g | | 2 (-6) |

This research work select the Koha open source library management software because it support federated search facilities by the Z39.50 server in admin interface. The Figure – 14 indicates the Z39.50 server which helps the data import from the other library OPAC. All the Z39.50 server informations are to be found from the irspyindex data websites and it will helps the college librarians to add the new Z39.50 server in developing the multilingual federating search system.

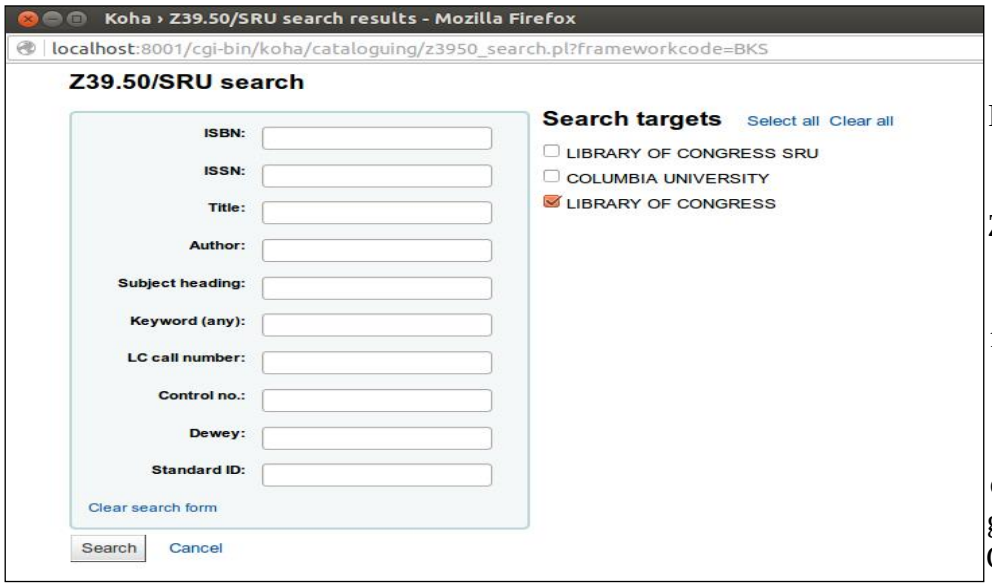


Figure – 14: Z39.50 in federated searching for CLBU

### 1.7.2 Mulitilingual Data Edit through MarcEditor for CLBU

MarcEditor is a data conversion tool and it supports the mulitilingual. Data migration is possible through this tool from one system to another. In previous most of the college libraries have been using the closed source software and local software including SOUL, Libsys, WINISIS and other local software. In these commercial software there is no standard to manintain and managed the bibliographic and authority data in the libraries. These research work successfully convert the MARC data both bibliographic as well as authority data by using the MarcEditor open source

tools. Mulitilingual can not managed through these non-standard based commercial software. Now, this problem can be solved through Koha open source relevant and popular library management software. It is also support the multilingual documents including Bengali, English, Hindi and etc. The multilingual data edit interface of MarcEditor is to be represents in the Figure – 15 for the college libraries in domain specific data conversion and data migration.

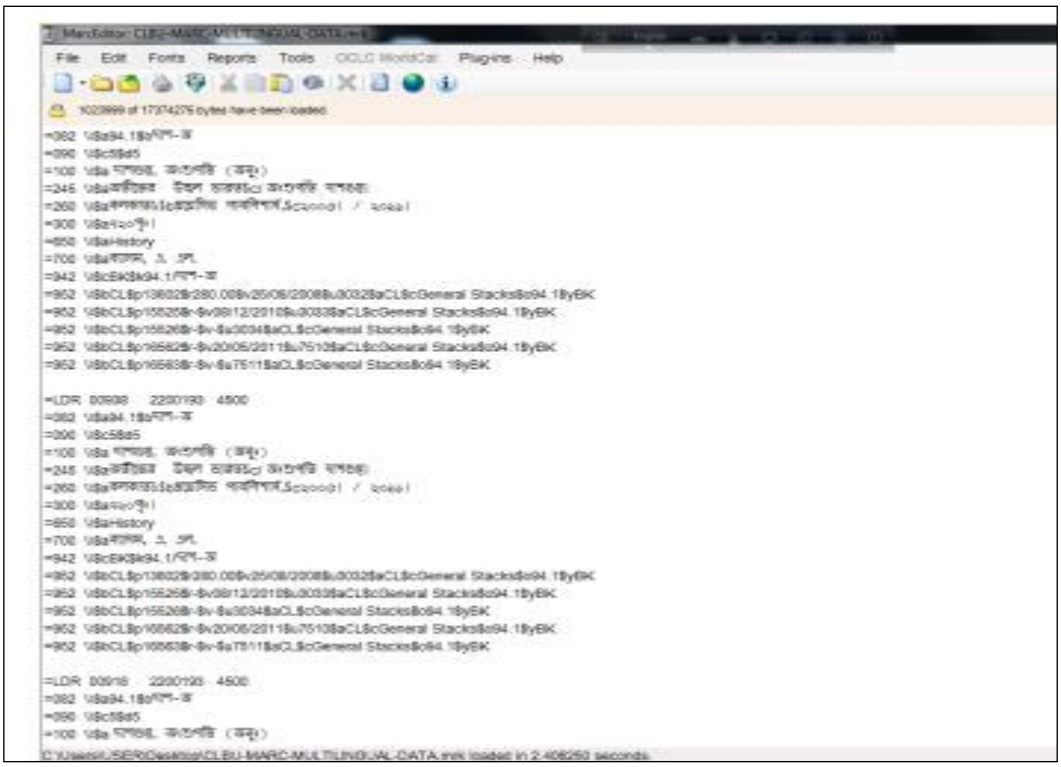


Figure – 15 : Multilingual data in MarcEditor for libraries

### 1.7.3 Content Creation and Editing in Mulitilingual for libraries

Joomla is the most popular and eminent software in the field of content creation in multilingual for the libraries. Required multilingual database connectivity through PhpMyAdmin because it support more languages. Backup and restorations is also possible through this tool in the domain specific cluster. Figure – 16 is the database connectivity software for the content as well as connect other softwares in six domain specific clusters including integrated library system cluster, learning content management system, community communication and interactions and also content management system (Rahimi, Shakery & King, 2015).

.

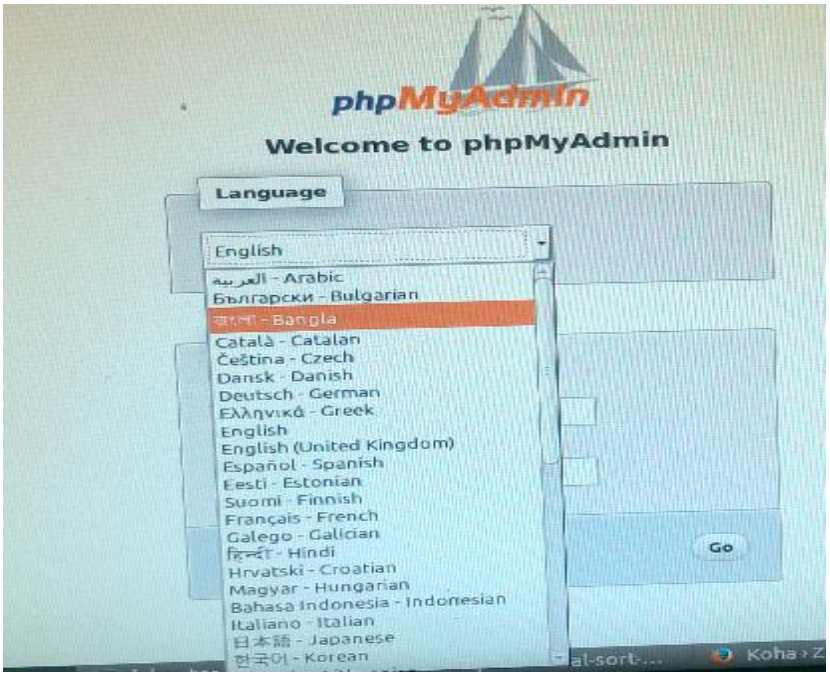


Figure – 16 : Multilingual in PhpMyAdmin for libraries

Joomla installation environment appear all the languages (Figure – 17) in the field of content creation, editing, deleting and also it helps the Website building for the college libraries. Joomla is the most popular and widely supported open source multilingual CMS platform in the world, offering more than 64 languages and by installing a language pack that will translate from the Joomla admin panel (Ashraf & Gulati, 2013). And after the users can go through some simple setting steps like getting in the content languages, language switcher, menus translated. It also built in capabilities to create a multilingual website. No additional plugins and components need to be installed in order to be able to translate your website. Multilingual image can be upload by using this software (Si & Callan, 2006).

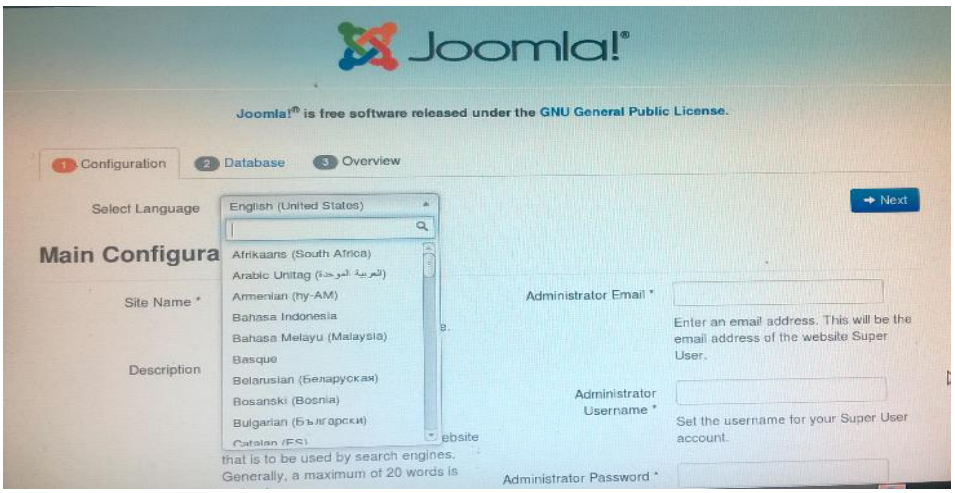


Figure – 17 : Multilingual in Joomla for libraries

Check out this demo installation of Joomla to see how multilingual websites work. Click on the country flags in the left-hand navigation (See Figure – 18) to see how websites look in different languages. The college librarians can do this by navigating to Extensions Manager -> Install Languages, selecting the language(s) you wish to install, and clicking the yellow Install button in the upper-right area of the page.

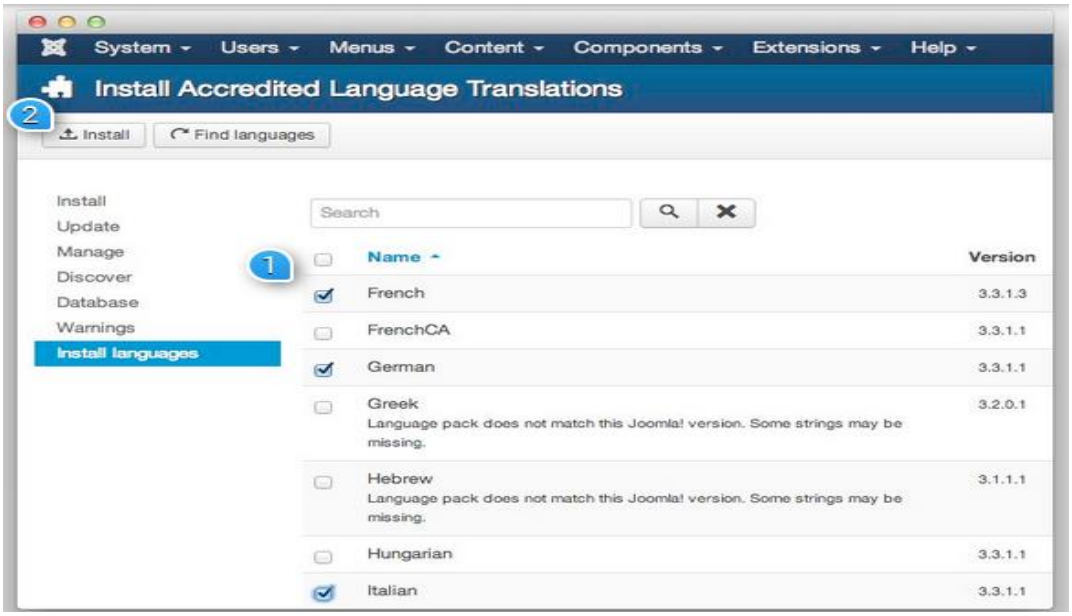


Figure – 18: Language installation in Joomla for CLBU

Setting up a basic Drupal website in English is relatively easy and also Setting up a multilingual website for the college libraries. Download, install and activate the i18n and Variable modules (and all their submodules) (Ruiz & Chin, 2010). The Variable module is new and required by i18n in D7. It provides a simple interface where you can designate system variables as Multilingual variables. and configure in the settings.php file. Translation dashboard in Drupal represents in the Figure – 19 for developing the multilingual content in the college libraries.

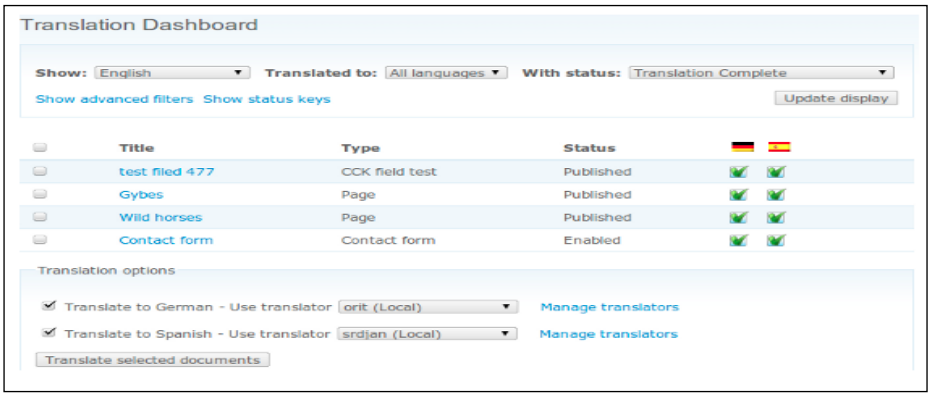


Figure – 19: Translation dashboard in Drupal

## 1.8 Findings and Conclusion

The findings of this paper for multilingual document management are as follows:

(i) It is support the Unicode base multilingual standards and components.

(ii) It is possible to development of indic script based retrieval system for libraries.

(iii) Google indic transliteration is also possible from the library OPAC in Koha.

(iv) Virtual keyboard integration is possible and its access from the Koha OPAC.

(v) Total twenty two languages are appeared in Koha OPAC for students and library professionals.

(vi) Installation of Bengali languages in Koha both the librarian and OPAC interfaces.

(vii) Setting up the multilingual from the Koha system administration.

(viii) Translate the Koha interface of each folders against in different modules and sub-modules also.

(ix) Changing the Koha interface in Bengali language.

(x) Data entry through Ibus avro phonetics in Koha for bibliographic records.

(xi) Retrieved Search results in Bengali language in Koha.

(xii) Development of multilingual in digital media archiving cluster by DSpace.

(xiii) Create the metadata in Bengali language from the DSpace digital library software.

(xiv) Mulitilingual is also managed in other clusters like content management system, learning content management system, community communication interaction and federated search system.

(xv) VuFind discovery tool can easily retrieved the multilingual library resources from the user interfaces.

(xvi) Import the mltilingual bibligraphic and authority data from the other library OPAC through the Z39.50 server.

As far as the libraries in the state of West Bengal are concerned, multilingual resource management is an important activity for college libraries as in some college libraries regional language based resources cover up to seventy percent of the collection. It shows the achievements of the said objective by develop mechanisms to managing, processing and retrieval of multilingual resources in Unicode - compliant environment including provisions for easy – to use input tools for different Indic - scripts with

special emphasis on Bengali script. This research work has integrated Avro - Phonetics and three other virtual keyboards in end user interfaces as well as in data entry interfaces. For example, the Google Indic transliteration facility with almost all Indic scripts (22 constitutionally recognized languages) is also made available in end user retrieval interfaces. It is quite easy to apprehend that the software framework with six domain - specific clusters and an array of open source tools for end users is very complex to implement at the user end.

**References**


Alshawi, H. (1992). The Core Language Engine . Cambridge, Ma: MIT Press.

Angelov, K. (2008). Type-Theoretical Bulgarian Grammar. In B. Nordstr¨om and A. Ranta (Eds.), *Advances in Natural Language Processing* (Go- TAL 2008) , Volume 5221 of LNCS/LNAI , pp. 52– 64. URL http://www.springerlink.com/ content/978-3-540-85286-5/ .

Angelov, K. (2009). Incremental Parsing with Parallel Multiple Context-Free Grammars. *In Proceedings of EACL'09*, Athens .

Angelov, K. and A. Ranta (2009). Implementing Controlled Languages in GF. *In Proceedings of CNL- 2009, Marettimo*, LNCS. to appear.

ANSI/NISO Z39.50 (1995). Information retrieval (Z39.50). Application service definition and protocol specication.

Ashraf, T., & Gulati, P. A. (2013). *Design, development, and management of resources for digital library services*. Hershey, Pa: IGI Global (701 E. Chocolate Avenue, Hershey, Pennsylvania, 17033, USA.

Bar-Hillel, Y. (1964). Language and Information . Reading, MA: Addison-Wesley.

Beckert, B., R. H¨ahnle, and P. H. Schmitt (Eds.) (2007). Verification of Object-Oriented Software: The KeY Approach . *LNCS 4334*. Springer-Verlag.

Bender, E. M. and D. Flickinger (2005). Rapid prototyping of scalable grammars: Towards modularity in extensions to a language-independent core. *In Proceedings of the 2nd International Joint Conference* on Natural Language Processing

Bergman, M. (2001). The deep web: ACM Press / Addison Wesley. Surfacing the hidden value. http://www.completeplanet.com/Tutorials/DeepWeb/index.asp. BrightPlanet.

Buchinski, E. J., Newman, W. L. & Dunn, M . J. (1976). The automated authority subsystem at the National Library of Canada. *Libr . Autom*, 9 (4): 279- 298.

Buttenfield, B. (1999). Usability evaluation of digital libraries. Digital libraries: philosophies, technical design considerations, and example scenarios, D. Stern (ed.), Binghampton, NY: Haworth Press, p.39-60.

Cook, V. (1999). Going beyond the native speaker in language teaching. *TESOL Quarterly*, 33 (2), 185-209.

Cox, C. N. (2007). *Federated search: Solution or setback for online library services*. Binghamton, NY: Haworth Information Press.

Craswell, N. & Hawking, D. (2000). Merging results from isolated search engines. *In Proceedings of the 10th Australasian Database Conference*. (pp. 189-200).

Fuhr, N. (2007). Evaluation of digital libraries. *International Journal on Digital libraries*, Springer, 18 p.

Gazen, B. and Minton, S. (2005). AutoFeed: an unsupervised learning system for generating webfeeds. *In Proceedings of the third International Conference* on

Knowledge Capture, ACM.
Janssen, Olaf (2003). Gabriel 1997-2003 & Gabriel/TEL user survey.
Liu. K. L., Yu. C., & Meng. W., Santos. A., & Zhang. C. (2001). Discovering the representative of a search engine. *In Proceedings of 10 th ACM International Conference* on Information and Knowledge Management (CIKM). ACM.
Mizera-Pietraszko, J., & Zgrzywa, A (2010). Vertical Search Strategy in Federated Environment.
Mudawwar, Muhammad F. (1997). Multicode: A Truly Multilingual Approach to Text Encoding. *Computer*, 30 (4), 37–43, April 1997.
Oakes, M., Xu, Y. (2009). A Search Engine based on Query Logs and Search Log Analysis at the University of Sunderland. In Peters, C. (Ed.), Results of the CLEF 2009 Cross-Language System.
Osborne, R., & American Library Association (2004). *From outreach to equity: Innovative models of library policy and practice*. Chicago: American Library Association.
Paolillo, J., Pimienta, D., Prado, D. (2007). Measuring linguistic diversity on the Internet. UNESCO.
Ponte, J. M & Croft, W. B. (1998). A language modeling approach to information retrieval. *In Proceedings of the 25 th Annual International ACM SIGIR Conference* on Research and Development in Information Retrieval. ACM. Publications for the World Summit on the Information Society. Retrieved from http://unesdoc.unesco.org/images/0014/001421/ 1421 86e.pdf
Peters, C., Braschler, M., Clough, P.D. (2012). Multilingual Information Retrieval: From Research to Practice. Berlin, Heidelberg: Springer.
Powell, J., & Fox, E. A. (1998). Multilingual Federated Searching Across Heterogeneous Collections. *D-lib Magazine.*
*Rahimi, R., Shakery, A., & King, I. (2015). Multilingual information retrieval in the language modeling framework. Information Retrieval Journal, 18, 3, 246-281.*
Ranta, A. (2004). Grammatical Framework: A Type-Theoretical Grammar Formalism. *The Journal*
*of Functional Programming*, 14 (2), 145– 189. URL http://www.cs.chalmers.se/ aarne/articles/gf-jfp.ps.gz.
Roberson, S. E. & Walker, S. (1994). Some simple effective approximations to the 2-Poisson model for probabilistic weighted retrieval. *In Proceedings of the 17th Annual International ACM SIGIR Conference* on Research and Development in Information Retrieval. ACM.
Rogati, M. & Yang, Y. (2003). CONTROL: CLEF-2003 with open, transparent resources off-line. experiments with merging strategies. In C. Peters(Ed.), Results of the CLEF2003. cross-language evaluation forum.
Rountree, D. (2012). *Federated Identity Primer*. Burlington: Elsevier Science. Cook, V. (1999). Going beyond the native speaker in language teaching. *TESOL Quarterly*, 33 (2), 185-209.
Ruiz, M. E., & Chin, P. (2010). Users' seeking behavior and multilingual image tags. *Proceedings of the American Society for Information Science and Technology, 47*, 1, 1-2.
Savoy, J. & Rasolofo, Y. (2000). Report on the TREC-9 experiment: link-based retrieval and distributed collections. *In Proceedings of 9th Text REtrieval Conference (TREC-9).* National Institute of Standards and Technology, special

publication 500-249.
Shokouhi, M., & Si, L. (2011). *Federated search*. Boston: Now Publishers.
Si, L., & Callan, J. (2006). CLEF 2005: Multilingual Retrieval by Combining Multiple Multilingual Ranked Lists.
Stiller, J., Gäde, M., Petras, V. (2013). Multilingual Access to Digital Libraries: The Europeana Use Case. *Information - Wissenschaft & Praxis*, 64 (2-3), 86 - 95.
Tran, D. T. (2011). *Process-oriented Semantic Web Search*. Amsterdam: IOS Press.
TrebleCLEF (2008). D3.2, Workshop on Best Practices for the Development of Multilingual Information Access Systems: the User Perspective. http://www.trebleclef.eu/getfile.php?id (Accessed on July 15, 2014).
Turtle, H. (1990). Inference networks for document retrieval. Technical Report COINS Report 90-7,
Computer and Information Science Department, *University of Massachusetts,* Amherst.
UNESCO (2003). Education in a multilingual world. Retrieved from http://unesdoc.unesco.org/ images/0012/001297/129728e.pdf. (Accessed on December 22, 2015)
Vijaya, MS; Ajith, VP; Shivapratap, G and Soman, KP (2009). English to Tamil Transliteration using WEKA, *International Journal of Recent Trends in Engineering*, 1 (1), 498-500.
Viles, C. L. & French, J. C. (1995). Dissemination of collection wide information in a distributed 157information retrieval system. *In Proceedings of the 18 th Annual International ACM SIGIR Conference* on Research and Development in Information Retrieval. ACM.
Voorhees, E., Gupta, N. K., & Johnson-Laird, B. (1995). Learning collection fusion strategies. *In Proceedings of the 18th Annual International ACM SIGIR Conference* on Research and Development in Information Retrieval . ACM.
Yang, Y. (1999). An evaluation of statistical approaches to text categorization. *Information Retrieval* . (1) (pp 69-90).
Yuwono, B. & Lee, D. L. (1997). Server ranking for distributed text retrieval systems on the Internet. *In Proceedings of the 5th Annual International Conference* on Database Systems for Advanced Applications. (pp. 41-49). World Scientific Press.
Zhai, C. X. & Lafferty, J. (2001). A study of smoothing methods for language models applied to ad hoc information retrieval. *In Proceedings of the 24 th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM.
Zhu, X. J. (2005). Semi-Supervised learning with graphs. Ph. D. Thesis, *Language Technology Institute, Carnegie Mellon University*.

**Additional Readings**

Brice, A. E. (2015). Multilingual Language Development. In J. D. Wright (Ed.), *International Encyclopedia of the Social & Behavioral Sciences (Second Edition)* (Second Edition, pp. 57 – 64). Oxford: Elsevier. Retrieved from http://www.Sciencedirect.com/science/article/ pii/B 9780080970868231267
Cruz, F. L., Troyano, J. A., Pontes, B., & Ortega, F. J. (2014). Building layered, multilingual sentiment lexicons at synset and lemma levels. *Expert Systems*

*with Applications*, *41*(13), 5984 – 5994. http://doi.org/http://dx.doi.org /10.1016 /j.es wa.2014.04.005

Granell, X. (2015). 9 - From {PLEs} to PLWEs: a Multilingual Information Management System. In X. Granell (Ed.), *Multilingual Information Management* (pp. 157 – 163). Oxford: Chandos Publishing. Retrieved from http://www.sciencedirect. com/science/article/pii/B97 81843347712000093

Ismaili, M. (2015). Teaching English in a Multilingual Setting. *Procedia - Social and Behavioral Sciences*, *199*, 189 – 195. http://doi.org/http://dx.doi. Org/10 .10 16 / j.sbspro.2015.07.505

Jonsson, C., & Muhonen, A. (2014). Multilingual repertoires and the relocalization of manga in digital media. *Discourse, Context & Media*, *4–5*, 87 – 100. http://doi.org/http://dx.doi.org /10.1016/j.dcm.2014.05.002

Singh, P. K., Sarkar, R., & Nasipuri, M. (2015). Offline Script Identification from multilingual Indic-script documents: A state-of-the-art. *Computer Science Review*, *15–16*, 1 – 28. http://doi.org/http://dx.doi.org/ 10.1016/j.cosrev. 2014 . 1 2 . 0